

Reviewed Preprint
v1 • July 2, 2026
Not revised

A membrane insertion code for intrinsically disordered proteins

✉ For correspondence:

hzhou43@uic.edu

Competing interests: No competing interests declared

Funding: See page 19

Reviewing editor: Heedeok Hong, Michigan State University, United States

© 2026, Muhammedkutty & Zhou. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Fidha Nazreen Kunnath Muhammedkutty¹, Huan-Xiang Zhou^{1,2} ✉

¹Department of Chemistry, University of Illinois Chicago, Chicago, United States, • ²Department of Physics, University of Illinois Chicago, Chicago, United States

eLife Assessment

This study provides a **valuable** advance in understanding how disordered proteins interact with cell membranes by identifying the sequence rules that enable aromatic residues to penetrate deeply into the membrane interior. The integration of complementary computational approaches, including molecular simulations, large-scale sequence analysis, and the development of an online prediction server, makes the work potentially impactful for the membrane protein and intrinsically disordered protein communities. The evidence supporting the main conclusions is generally **convincing**, although its transferability across diverse membrane compositions and its validity as a prediction tool for real protein-membrane systems remain to be further established.

<https://doi.org/10.7554/eLife.111515.1.sa4>

Abstract

Membrane association of intrinsically disordered proteins (IDPs) mediates various cellular functions including membrane remodeling and signal transduction. Whereas membrane association through amphipathic helices and polybasic motifs is well understood, sequence determinants for deep membrane insertion of aromatic residues are still poorly characterized. Here, we decipher the sequence code for membrane insertion of aromatic-centered motifs. For an initial set of 10 9-residue aromatic-centered sequences, all-atom molecular dynamics simulations and the positioning of proteins in membranes (PPM) method produced very similar membrane insertion propensities. Applying PPM to a full library of 1.2×10^6 sequences with an F, W, or Y residue flanked by L, R, G, N, or E at four positions on either side, we found that aliphatic (L) and basic (R) residues favor membrane insertion, whereas acidic (E) and polar (N) residues disfavor it. Guided by these rules, we developed a mathematical model dubbed AroMIP to predict the membrane insertion propensities of aromatic-centered motifs. AroMIP achieves 91.2%, 92.0%, and 99.7% accuracies for F-, W-, and Y-centered motifs, respectively, in disordered regions of the human proteome and is available as a web server at <https://zhougroup-uic.github.io/AroMIP/>. The present work provides the sequence basis and a mechanistic understanding of how IDPs employ aromatic-centered motifs to drive membrane insertion, and enriches the tools for the study of IDP-membrane association.

Significance Statement

Membrane insertion of short motifs, along with membrane tethering of amphipathic helices and membrane binding of polybasic motifs, is a major mode of membrane association and mediates diverse functions including membrane remodeling and signal transduction. Here we used three complementary approaches to decipher the sequence code of membrane insertion, culminating in a sequence-based method, AroMIP, for predicting membrane-insertion propensities. Aromatic sidechains have the intrinsic ability to insert deeply into the acyl chain region of membranes; at flanking positions, they strongly stabilize the inserted state. Aliphatic and basic residues are

medium to modest promoters of membrane insertion. AroMIP has >90% accuracy and identifies important motifs for regulating functions of intrinsically disordered proteins via membrane insertion.

Introduction

Membrane association of intrinsically disordered proteins or regions (IDPs / IDRs) mediates a host of cellular functions, including membrane remodeling,^{1–6} ion channel gating,^{7–9} signal transduction,^{10–17} membrane anchoring,^{18, 19} and host-pathogen interactions.^{17, 20} A proteome-wide screening of human transmembrane proteins found ~60% of them to contain membrane-binding IDRs.²¹ IDRs can associate with membranes via at least three distinct modes: amphipathic helices, polybasic motifs, and aromatic-centered short motifs.²² Amphipathic helices most likely form upon membrane contact, resulting in a nonpolar face that projects into the acyl chain region and a polar / charged face that projects into the headgroup region or the aqueous phase. Polybasic motifs bind to the membrane surface, in particular via electrostatic interactions between R and K sidechains and phosphate groups of lipids (especially acidic lipids). Methods have been developed to predict IDP sequences that likely associate with membranes through these two modes.^{7, 23} In comparison, sequence determinants for deep membrane insertion of aromatic-centered short motifs are still poorly defined.

Aromatic residues are known to favor the membrane-water interface, based on the free energies of transfer from interface to water,²⁴ the distribution of amino acids in transmembrane proteins at different positions along the membrane normal,²⁵ and the contributions of interfacially positioned aromatic residues to the folding stability of β -barrel transmembrane proteins.^{26, 27} In the interface, aromatic sidechains can form cation- π interactions and hydrogen bonds (in the case of W and Y) with lipid headgroups and possibly hydrophobic interactions with acyl chains.^{28–32} F sidechains are also frequently positioned deeply toward the bilayer center.^{25, 26} Whereas the positions of residues in a transmembrane protein relative to the bilayer center are dictated by the three-dimensional structure of the protein, the counterparts in a peripheral membrane protein can be varied by a change in the burial depth of a local element (e.g., a loop) or an entire domain. Aromatic residues are enriched in the membrane-binding faces of peripheral membrane proteins;³³ whether these residues insert into the acyl chain region may be context-dependent. For example, based on the effects of fluorination on membrane binding, F249 on a loop of phosphatidylinositol-specific phospholipase C penetrates into the acyl chain region, but Y258 introduced on a neighboring helix is located in the headgroup region.³⁴ Cryo-electron microscopy (cryo-EM) structures of OPA1, a dynamin superfamily protein mediating fusion of the inner mitochondrial membrane, show that W775 and W778 in the flexible membrane-inserting loop (residues 768–782) are positioned deeply into the acyl chain region of cardiolipin-containing membranes.^{35, 36} Likewise, cryo-EM structures and all-atom molecular dynamics (MD) simulations suggest penetration into the acyl chain region by aromatic residues at the tip of the β 1– β 2 loop in gasdermins, pore-forming proteins that cause membrane permeabilization and pyroptosis (a pro-inflammatory lytic cell death).^{37–39} All-atom MD simulations also showed that W41 in a loop in the phloem lipid-associated family protein PLAT1 transiently inserts into the acyl chain region.⁴⁰

Aromatic residues are also enriched in membrane-binding IDRs²¹ and frequently noted in the membrane binding or insertion of IDPs. For example, using magic-angle spinning ^1H - ^1H nuclear Overhauser enhancement spectroscopy (NOESY), Zhang et al.¹⁰ observed strong cross peaks between acyl chains in DMPC:DMPG lipid membranes and five phenol rings from a 25-residue fragment (residues 152–176) of the IDP myristoylated alanine-rich C kinase substrate (MARCKS). Similarly, solution ^1H - ^1H NOESY of HIV-1 trans-activator of transcription (Tat), an 86-residue IDP, in DPC micelles revealed cross peaks between acyl chains of DPC and the indole ring of the lone W residue, W11.⁴¹ Based on chemical shift perturbation and NMR intensity loss due to line broadening upon binding cardiolipin-containing nanodiscs, Mahajan et al.⁵ identified both an amphipathic helix and an unstructured segment (T548–A566) in the variable domain (residues 497–607) of dynamin-related protein 1 (Drp1) for membrane binding. These authors further proposed that W552 in the latter segment interacts with acyl chains, based on analogy with a WRG sequence

in a transmembrane domain. A recent cryo-EM structure of Drp1 suggests membrane insertion of the variable domain.⁶ Loss of NMR intensity upon binding DMPC:DHPC or DMPG:DHPC bicelles also revealed a Y511-centered motif (and a Y483-centered motif to a lesser extent) in the pathogen-encoded signaling receptor Tir for membrane binding.¹⁷ It is unclear whether these Y sidechains insert into the acyl chain region. Based on paramagnetic relaxation enhancement by membrane-embedded spin labels, Jin et al.¹⁵ estimated that two Y residues, Y188 and Y199, in the immunoreceptor tyrosine-based activation motif (ITAM) (two Yxx(L/I) segments separated by six to eight residues) of the CD3 ϵ component of the T-cell receptor complex had 15% to 34% probabilities of being membrane-bound. Moreover, Y199 had ~20% probability of reaching the acyl chain region. These results stand in contrast to a static structure showing deep insertion of both Y residues solved by solution NMR spectroscopy.⁴² Lipid binding assays of fragments covering the cytoplasmic IDR of the ζ component showed that basic clusters outside the 3 ITAMs, instead of the ITAMs, bind to acidic membranes.¹¹

Coarse-grained MD simulations of the T-cell receptor complex embedded in a membrane showed that Y199 of CD3 ϵ had a propensity to penetrate into the acyl chain region; the propensity was lower for Y188 and even lower for Y residues in the 3 ITAMs of ζ chains.⁴³ Also using coarse-grained simulations, Araya-Secchi et al.¹⁴ observed that the I₂₆₇FPV₂₇₁ motif in the cytoplasmic IDR of the prolactin receptor interacted with acyl chains more often than with headgroups. Other aromatic-centered motifs were found to penetrate into the acyl chain region in all-atom simulations. One end of the spectrum is stable insertion, represented by the W₄₄₀FFGW₄₄₄ motif in a 57-residue IDR of Bap1, which mediates adhesion of *Vibrio cholerae* biofilms to host membranes.²⁰ The other end of the spectrum is brief excursion, e.g., by Y853 in the cytoplasmic IDR of metabotropic glutamate receptor 3 (mGluR3).¹⁶ Somewhere in between are the F₄₄₈GI₄₅₀ motif in the disordered juxtamembrane domain (residues 440-465) of tropomyosin receptor kinase A (TrkA),¹³ V₉F₁₀ in the N-terminal IDR of the bacterial K⁺ channel KtrB,⁹ and W21, W25, and W29 in the N-terminal IDR of ABHD5, a regulator of lipases.⁴⁴ Some of these simulation results have experimental support. For example, an F-to-S mutation in the F₄₄₈GI₄₅₀ motif weakened the binding of the TrkA juxtamembrane domain to acidic membrane, whereas an R-to-W mutation at position 452 strengthened membrane binding.¹³ For Bap1, a blue shift in tryptophan fluorescence upon membrane binding is consistent with a change from an aqueous environment to a hydrophobic environment for W sidechains, thereby implicating burial in the acyl chain region.²⁰ Similarly, membrane binding protects W21, W25, and W29 in ABHD5 from deuterium exchange, again implicating burial.⁴⁴ The Y853F mutation in mGluR3, introduced to deepen membrane penetration, indeed strengthened the membrane binding of residues 845-860.¹⁶

Determining whether a specific residue in an IDP inserts into the acyl chain region presents a major technical challenge. Experimentally, the most useful technique is NMR spectroscopy. Although both solid-state and solution ¹H-¹H NOESY have identified membrane-inserted residues,^{10, 41} both face difficult issues including resolution and sensitivity. All-atom MD simulations can map out the equilibrium distributions of IDP residues with respect to the membrane surface as well as the dynamic process of membrane insertion, such simulations are always open to the question of whether sampling is sufficient.⁴⁵ The accuracy of force fields is another issue, though more so for coarse-grained simulations.

Here we address this challenge by developing a fast, sequence-based method for predicting whether aromatic-centered motifs of IDPs insert into the acyl chain region of membranes. We focus on aromatic amino acids because, as made clear by the foregoing paragraphs, they are the most frequent ones among the 20 types of amino acids found participating in deep membrane insertion. We first performed extensive all-atom MD simulations of 10 aromatic-centered peptides (9 residues each) near POPC:POPS:PIP₂ (70:25:5 molar ratio) membranes to characterize their membrane insertion propensities. For high throughput, we used the positioning of proteins in membranes (PPM) method⁴⁶ and obtained membrane insertion propensities very similar to those from MD simulations. We then applied PPM to an exhaustive sequence library where each of the four flanking positions on either side of an aromatic residue was filled with one of five amino acids: L, R, G, N, and E. The PPM output was cell captured by a mathematical model, dubbed

AroMIP, in which each flanking residue contributed a multiplicative factor to the membrane insertion score. Lastly, we trained and tested AroMIP on IDRs in the human proteome,⁴⁷ achieving 91.2%, 92.0%, and 99.7% accuracies for F-, W-, and Y-centered motifs, respectively. AroMIP is available as a web server at <https://zhougroup-uic.github.io/AroMIP/>.

Results

All-atom MD simulations characterize membrane-bound states of aromatic-centered peptides and pathways leading to membrane insertion

We first collected 10 aromatic-centered motifs that previous studies have classified as either membrane inserters or non-inserters (Fig. 1A). This initial set includes an F-centered motif from TrkA (sequence 1), which inserted deeply into the acyl chain region in MD simulations;¹³ a W-centered motif in Drp1 (sequence 2), which was indicated by solution NMR to be an inserter;⁵ two Y-centered segments composing the ITAM in CD3 ϵ (sequences 3 and 4), which were estimated to have 20% and 0% probabilities, respectively, of reaching the acyl chain region based on modeling paramagnetic relaxation enhancement data;¹⁵ and six Y-centered segments composing three ITAMs in CD3 ζ (sequences 5 to 10), which did not bind to membranes according to binding assays.¹¹

For each sequence, we generated an initial disordered conformation using TraDES⁴⁸ and placed it near a membrane, with the central aromatic residue just above the phosphate plane. The peptide-proximal leaflet was composed of 70% POPC (zwitterionic), 25% POPS (net charge -1), and 5% PIP₂ (net charge -4). MD simulations were conducted in 20 replicates, each for 1 μ s. Figure 1B presents a simulation where the central aromatic sidechain (sequence 1), after \sim 90 ns, is stably inserted into the acyl chain region. We use a value of -3.1 \AA (indicated by a red dashed line), corresponding to the mean position of glycerol oxygen atoms, for the Z coordinate of the aromatic tip carbon atom (Z_{tip}) to demarcate insertion into the acyl chain region. In contrast, Fig. 1C presents a simulation where the peptide (sequence 4) moves away from the membrane.

We label a simulation as membrane-inserting if $>80\%$ of the frames in the last 100 ns have $Z_{\text{tip}} \leq -3.1$ \AA . Using this cutoff, the F- and W-centered peptides (sequences 1 and 2) have 3 and 8 simulations, respectively, out of a total of 20 each, as membrane-inserting (Fig. 1D). Among all their respective 20 simulations, the F- and W-centered peptides have 20.6% and 56.6%, respectively, membrane-inserting frames (Table 1). By contrast, none of the 20 simulations of each of the six Y-centered peptides (sequences 3-10) is membrane-inserting. We classify sequences 1 and 2 as membrane inserters (with at least one membrane-inserting simulation) whereas sequences 3-10 as non-inserters. Sequence 3 is special among the non-inserters: in two of its simulations, the percentage of membrane-inserting frames is substantial, at 57.0% and 22.4%, though not reaching the 80% threshold (to be called “partly inserting”). Correspondingly, the overall percentage of membrane-inserting frames in 20 simulations of this peptide is the highest, at 4.5%, among all six Y-centered peptides (Table 1). These outcomes are in complete agreement with the previous characterizations of the membrane-insertion propensities of these peptides.^{5, 11, 13, 15}

We further analyzed the simulations of sequences 1-3 to characterize membrane-bound states and pathways leading to membrane insertion. For both the F- and W-centered peptides, the aromatic Z_{tip} distributions in the last 100 ns of all 20 simulations feature a peak at -6.5 \AA (corresponding to the inserted state; Fig. 2A, B), where the aromatic sidechains form hydrophobic interactions with acyl chains. According to the tilt angle of the aromatic ring (Fig. S1A), the inserted state splits into two populations (Fig. S1B, C). The major population has the aromatic ring close to parallel to the membrane normal (tilt angle \sim 150°), surrounded by acyl chains (Fig. S1B, C, snapshots on the right). In the minor population, the aromatic ring turns sideways (tilt angle \sim 90°), with the surrounding acyl chains also bent to maintain hydrophobic contact (Fig. S1B, C, snapshots on the left). In both populations, the indole nitrogen atom of W can hydrogen bond with

Fig. 1. Classification of aromatic-centered motifs into membrane inserters and non-inserters by MD simulations.

(A) Initial dataset containing 10 sequences (sequence 1 from TrkA, 2 from Drp1, 3 and 4 from CD3ε, and 5 to 10 from CD3ζ). (B) Initial snapshot, time trace of aromatic Z_{tip} , and snapshot at 980.7 ns for sequence 1. (C) Initial snapshot, time trace of aromatic Z_{tip} , and snapshot at 964.3 ns for sequence 6. (B) and (C) represent simulations where the peptides are inserted or not inserted, respectively, into the membrane. Insertion means at least 80% of the frames in the last 100 ns (shaded region) have aromatic Z_{tip} at least 3.1 Å (indicated by a red dashed line) below the phosphate plane. Time traces are smoothed using the moving average in a 2.5-ns window. The membrane is displayed as surface, with phosphates and all atoms above in orange while those below in gray; the peptide is displayed as cartoon in cyan, with the central aromatic residue as sticks in magenta and red. (D) The number of replicate simulations, out of a total of 20, where the peptide is inserted into the membrane.

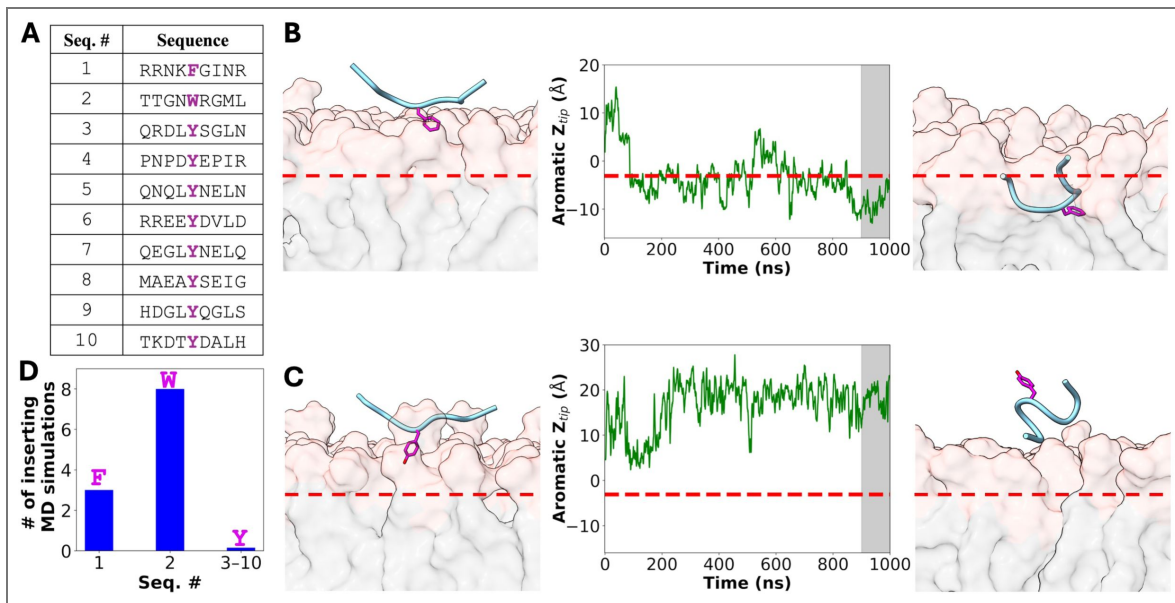


Table 1. MD, PPM, and AroMIP results on the initial dataset

Seq. #	MD % inserting frames	PPM % inserting conformations	AroMIP		
			Q	P	Inserter?
1	20.6	14.8	2.319	0.699	Y
2	56.6	47.6	3.591	0.782	Y
3	4.5	0.5	0.011	0.011	N
4	0.0	1.7	0.038	0.037	N
5	1.2	0.1	0.004	0.004	N
6	0.0	0.0	0.012	0.012	N
7	0.0	0.1	0.006	0.006	N
8	0.1	1.0	0.010	0.010	N
9	0.0	0.7	0.002	0.002	N
10	0.0	0.1	0.002	0.002	N

glycerol oxygen atoms. Another difference between F-centered sequence 1 and W-centered sequence 2 is that the latter has a second peak at $Z_{\text{tip}} = -2.5 \text{ \AA}$, in the headgroup region, whereas the former has a second peak at $Z_{\text{tip}} = 7.5 \text{ \AA}$, above the membrane. In comparison, for the Y-centered sequence 3, the peak with the lowest Z_{tip} is at -2.5 \AA (above the cutoff for membrane insertion and termed partially inserted), where the sidechain hydroxyl can hydrogen bond with glycerol oxygen atoms (Fig. 2C). In the subpopulation where the aromatic ring is inserted into the acyl chain (with tilt angle $\sim 130^\circ$), a neighboring lipid is also lowered, such that the hydroxyl-glycerol hydrogen bond is maintained (Fig. S1D, E). Indeed, the Y sidechain and the glycerol sn-2 carbon atom of the partner lipid rise and fall together (Fig. S1E), with a strong correlation between the respective Z_{tip} and Z coordinate (correlation coefficient at 0.78; Fig. S1F). The Y-centered sequence 3 also has a second peak, at $Z_{\text{tip}} = 2.5 \text{ \AA}$ (Fig. 2C). For each peptide, the second peak in aromatic Z_{tip} constitutes an intermediate state.

Figure 2D compares the poses of aromatic sidechains in sequences 1-3 in their respective (partially) inserted states. In the last 100 ns of the membrane-inserting simulations of sequences 1 and 2, the F and W sidechains sample only the inserted state (Fig. S2A, B, red curves). In other, partly inserting simulations, the F and W sidechains sample both the intermediate and inserted states (Fig. S2A, B, insets). Of the two partly inserting simulations of Y-centered sequence 3, the aromatic sidechain is confined to the partially inserted state in the simulation with more inserting frames (57.0%) but samples both the intermediate and partially inserted states in the other simulation (Fig. S2C, solid and dashed red curves). At their respective intermediate states, all three aromatic sidechains have a tendency to form cation- π interactions with a lipid choline group (Fig. 2A-C). Cation- π interactions are formed in 13.8%, 26.4%, and 24.2% of all the membrane-interacting frames in the 20 MD simulations of sequences 1 to 3, respectively (Fig. S2).

Membrane insertion is not an act of the central aromatic sidechain alone. Rather, it is helped or hindered by the flanking residues. Figure S3A, B displays the Z_{tip} coordinates of all sidechains in F-centered sequence 1 and W-centered sequence 2 in the last 100 ns of the inserted simulations; Fig. S3C displays the corresponding results for Y-centered sequence 3 in the two partly inserting simulations. For all three sequences, the central aromatic sidechain is consistently inserted deeper than flanking sidechains, as shown by the average over replicate simulations (Fig. S3, solid black curves), justifying our focus on aromatic residues. For F-centered sequence 1 (Fig. S3A), the inserted state is stabilized by intermittent penetration of I[+2] (two residues upstream of the central aromatic residue) into the acyl chain region and snorkeling of R[+4], which means that its aliphatic portion contacts acyl chains while its terminal guanidinium group interacts with phosphates. Additional stabilization is provided by salt-bridge formation with lipid phosphates by all the basic sidechains. In the case of W-centered sequence 2 (Fig. S3B), membrane penetration by M[+3] and L[+4] and snorkeling and salt-bridge formation by R[+1] are the major stabilizing factors for the inserted state. For Y-centered sequence 3 (Fig. S3C), membrane penetration by L[+3] and salt-bridge formation by R[-3] reinforce membrane binding.

Inspecting the entire trajectories of sequences 1-3 (illustrated by Fig. 1B), we arrive at the following pathways for membrane insertion of aromatic-centered motifs. The pathways generally consist of three steps: initial membrane contact, further stabilization at the membrane surface, and insertion by the central aromatic sidechain (Fig. 3). These steps lead to poses illustrated and labeled as 1, 2, and 3 in circles in Fig. 3; poses 2 and 3 correspond to the intermediate and (partially) inserted states, respectively, in Fig. 2A-C. Pose 3 is already described in the preceding paragraph; we now describe poses 1 and 2 (Fig. S4). For F-centered sequence 1, R[-4], R[-3], and K[-1] make the initial membrane contact, via long-range electrostatic interactions with phosphates, especially those on acidic lipids (Fig. S4A; Fig. 3A pose 1). R[+4] then tethers the opposite end of the peptide to the membrane, priming the F sidechain for cation- π interaction with a choline (Fig. 3A pose 2). For W-centered sequence 2, R[+1] makes the initial membrane contact, followed by membrane penetration by L[+4] (or M[+3]) (Fig. S4B; Fig. 3B pose 1). With deeper penetration by L[+4] and M[+3], the W sidechain now forms cation- π interaction in the headgroup region (Fig. 3B pose 2). For Y-centered sequence 3, the initial membrane contact can occur

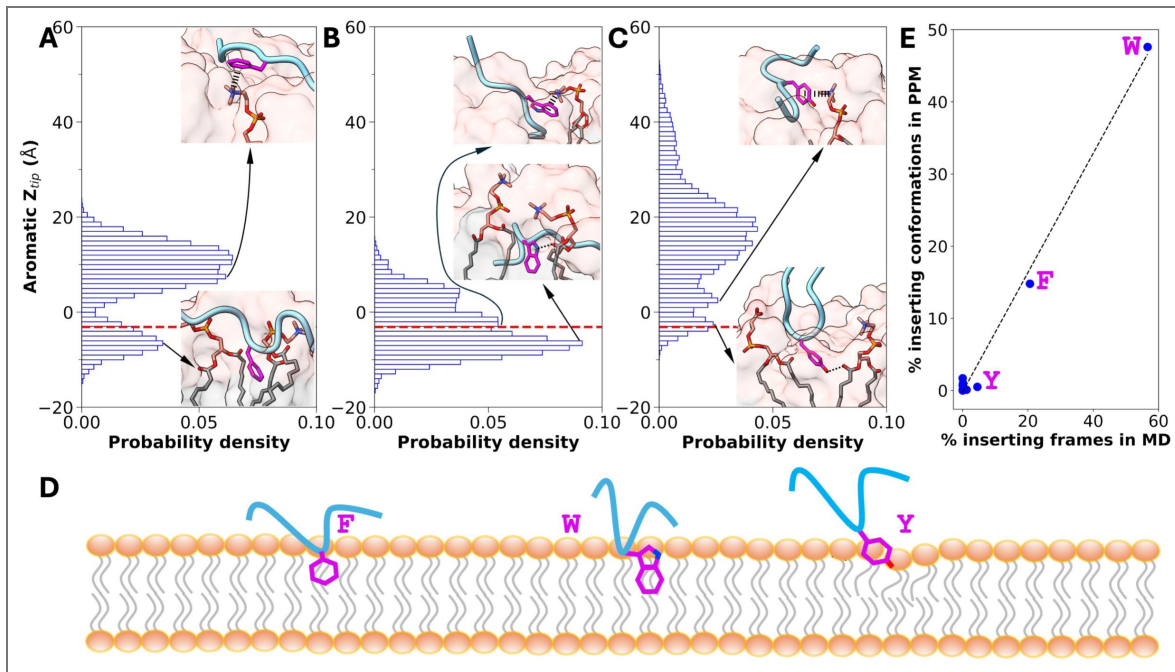


Fig. 2. Differences and similarities in membrane insertion and cation- π interactions among F-, W-, and Y-centered peptides.

(A) Aromatic Z_{tip} distribution for an F-centered peptide (sequence 1). Insets: snapshot with F deeply inserted to interact with lipid acyl chains (bottom), or above the membrane to form cation- π interaction with a choline group (top). (B) Aromatic Z_{tip} distribution for a W-centered peptide (sequence 2). Insets: snapshot with W deeply inserted to interact with lipid acyl chains and form a hydrogen bond with a glycerol oxygen atom (bottom), or in the headgroup region to form cation- π interaction with a choline group (top). (C) Aromatic Z_{tip} distribution for a Y-centered peptide (sequence 3). Insets: snapshot with Y in the headgroup region to form a hydrogen bond with a glycerol oxygen atom (bottom), or near the membrane to form cation- π interaction with a choline group (top). In (A-C), results were accumulated from the last 100 ns of 20 simulations of each peptide; a red dashed line is drawn at $Z_{tip} = -3.1 \text{ \AA}$. (D) Illustration of stable positions of F, W, and Y with respect to the membrane. (E) Extremely strong correlation between the percentage of membrane-inserting frames in MD simulations and the percentage of membrane-inserting conformations in PPM runs (correlation coefficient = 0.995).

through R[-3] via electrostatic interaction or the Y sidechain itself via cation- π interaction (Fig. S4C [↗](#); Fig. 3C [↗](#) pose 1). Then, R[-3] and L[+3] pin the peptide to the membrane, placing the Y sidechain for cation- π interaction with a choline (Fig. 3C [↗](#) pose 2).

Previous MD and NMR studies using longer constructs provide validation of our all-atom simulations using 9-residue fragments. Our F-centered sequence 1 was taken from the disordered juxtamembrane domain (residues 440-465) of TrkA; Wang et al.¹³ used all-atom MD simulations to characterize the insertion of this 26-residue region into membranes with a simplified representation. The insertion distance profile in that study is quantitatively reproduced by our simulations (Fig. S5 [↗](#)), both showing the deepest insertion by the central F residue. Likewise, our W-centered sequence 2 was taken from the disordered variable domain (residues 497-607) of Drp1. Solution NMR data of Mahajan et al.⁵ on this 111-residue region indicated that residues W₅₅₂RG are partially helical and insert into cardiolipin-containing membranes. In one simulation of sequence 2, WRGML form an α -helix lying parallel to the membrane surface (Fig. S6A [↗](#)), much like an amphipathic helix. In other membrane-inserting simulations, sequence 2 does not form an α -helix (Fig. S6B, C [↗](#)), showing that the helical structure is only partially stable. The amphipathic orientation in our simulation contrasts with a transmembrane orientation proposed by Mahajan et al. based on analogy with a transmembrane domain. Still, when averaged between simulations with helical and non-helical structures, the secondary chemical shifts have similar magnitudes (0.2 to 2 ppm; Fig. S6D [↗](#)) as reported by Mahajan et al.

PPM provides a fast alternative to all-atom MD simulations in separating inserters and non-inserters

As demonstrated above, all-atom MD simulations are very effective in separating membrane inserters from non-inserters and providing mechanistic insights, but they are time-consuming and therefore not feasible to apply to a very large set of sequences in order to reach a comprehensive understanding on the contributions of flanking residues to membrane insertion. To address this limitation, we developed a fast workflow based on the PPM method,⁴⁶ originally developed for membrane placement of structured proteins. We adapted this method for disordered peptides, by using a batch of random conformations generated by the TraDES algorithm⁴⁸ to represent each peptide (Fig. 4A [↗](#)). Each PPM run, using a peptide conformation as input, produced a pose of the peptide with respect to the membrane (Fig. 4B [↗](#)) and a membrane interaction free energy, ΔG . Figure 4C [↗](#) displays a scatter plot of ΔG and aromatic Z_{tip} data for a batch of 100 conformations of the F-centered sequence 1. We choose a cutoff of -1.5 Å to label a peptide conformation as membrane-inserting. This cutoff is less strict than in the all-atom MD simulations, due to the rigid treatment of peptides and implicit treatment of the membrane by PPM, but we compensate by adding a cutoff on ΔG , at -4.1 kcal/mol (Fig. 4C [↗](#), region bordered by red dashed lines).

We measure the membrane-insertion status of each peptide by the percentage of membrane-inserting conformations in a batch of 100 conformations, and denote this quantity as Φ . Replicate PPM runs yield highly reproducible results for each of the 10 sequences. The minimum Φ values of sequences 1 and 2 is 11% and 38%, respectively, while the maximum Φ value of sequences 3 to 10 is 3%. Figure S7 [↗](#) displays the mean Φ values and standard deviations for the 10 sequences, showing a clear gap between inserters (sequences 1 and 2) and non-inserters (sequences 3 to 10). Therefore, PPM can unequivocally separate inserters from non-inserters, just like all-atom MD simulations. Indeed, the mean percentage of inserting conformations in PPM (Table 1 [↗](#)) shows a very strong correlation with both the percentage of inserting frames in MD simulations (Fig. 2E [↗](#)) and the percentage of MD simulations labeled as membrane-inserting (Fig. 4D [↗](#)). In both cases, the correlation coefficients are very close to 1. This strong correlation firmly supports our use of PPM for fast classification of aromatic-centered motifs into membrane inserters or non-inserters.

To complete the PPM-based workflow, we had to choose a threshold Φ value, Φ_{th} , for separating each peptide sequence into inserters and non-inserters. We took into consideration the following factors. First, PPM did not account for lipid composition.⁴⁶ Cell membranes contain various levels of acidic lipids, which tend to attract basic sidechains for membrane association,^{7, 16} thereby facilitating membrane insertion of adjacent aromatic residues,^{9, 10} as also demonstrated in our all-

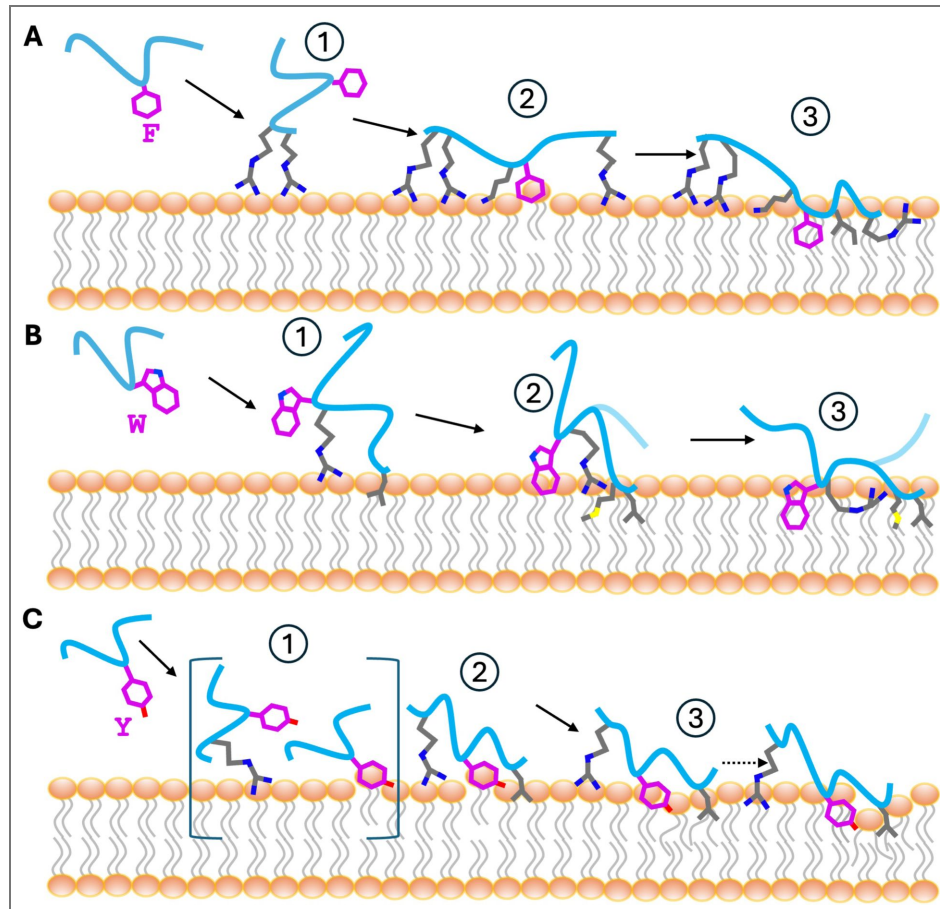


Fig. 3. Three-step pathways to membrane insertion.

(A) F-centered sequence 1. Basic sidechains at positions -4, -3, and -1 first tether one end of the peptide to the membrane surface. R[+4] then tethers the peptide at the other end, situating the central F for cation- π interaction. Finally, the aromatic sidechain is inserted into the acyl chain region. (B) W-centered sequence 2. The R[+1] sidechain initiates membrane contact. With possible stabilization by the membrane penetration of an aliphatic residue at position +3 or +4, the central W forms cation- π interaction with a choline and then enters the acyl chain region. (C) Y-centered sequences. The initial membrane contact can be formed by either a flanking basic sidechain (R[-3]; via electrostatic interaction) or the central Y itself (via cation- π interaction). After stabilizing at the membrane surface by R[-3] and L[+3], the central Y lowers into the headgroup region to form a hydrogen bond with a glycerol oxygen atom. The aromatic ring occasionally dips into the acyl chain region, bringing the lipid glycerol with it and producing local curvature.

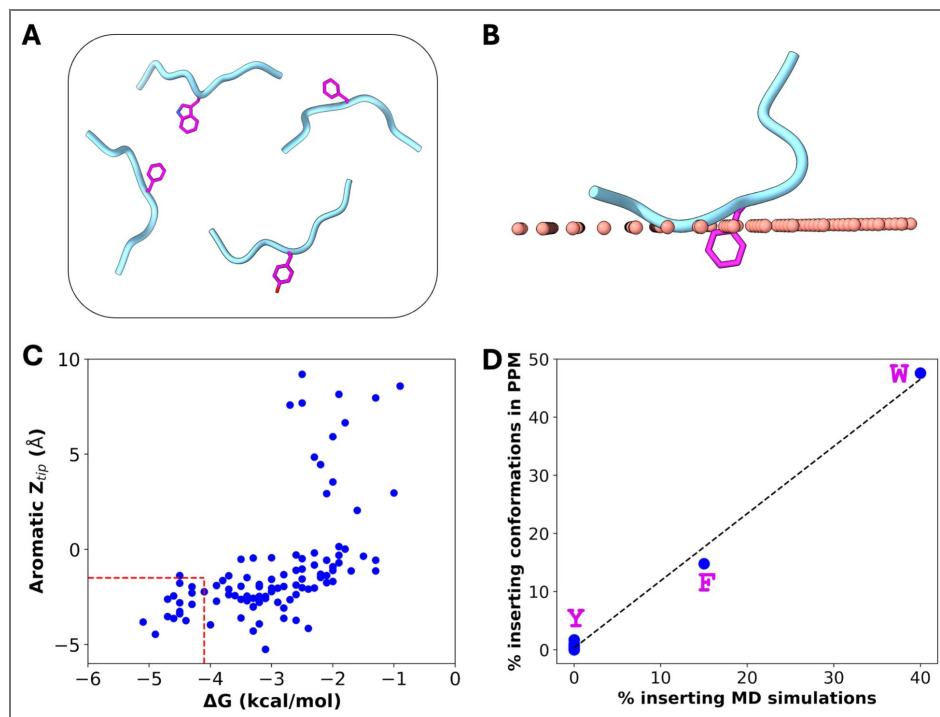


Fig. 4. Classification of aromatic-centered motifs into membrane inserters and non-inserters by PPM runs.

(A) Disordered conformations of 9-residue motifs generated by TraDES. (B) The pose of an F-centered peptide (sequence 1) from a PPM run. The phosphate plane is represented by arrays of small salmon spheres. (C) A scatter plot of ΔG and Z_{tip} data for a batch of 100 conformations of sequence 1. Dash lines represent cutoffs for defining inserters ($\Delta G < -4.1$ kcal/mol and $Z_{tip} < -1.5$ Å). 15 conformations satisfy the cutoffs. (D) Extremely strong correlation between the percentage of MD simulations classified as membrane-inserting and the percentage of conformations in PPM runs (correlation coefficient = 0.997).

atom MD simulations (Fig. 3A, B). We accounted for the contributions of flanking basic sidechains by lowering Φ_{th} . Second, Φ_{th} should fall in the gap between known inserters and non-inserters (Fig. S7). The end result for Φ_{th} is $40 - 15 \times R - 10 \times K$ (or 9 if that value is < 9), where R and K are the number of Arg and Lys residues, respectively, in the eight flanking positions of aromatic-centered motif. As a minimum test, applying this threshold to the PPM results (Table 1), sequences 1 and 2 are correctly classified as inserters and sequences 3-10 are correctly classified as non-inserters.

Expansion to a library of 1.2×10^6 random sequences reveals a membrane insertion code

We next applied the PPM-based workflow to an exhaustive library of sequences where the central position was filled with F, W, or Y and each of the four flanking positions on either side was filled with a reduced amino-acid alphabet comprising L, R, G, N, and E (Fig. 5A). This alphabet was selected to represent the chemical diversity of the 20 natural amino acids: L (Leu) as an aliphatic residue, R (Arg) as a basic residue, G (Gly) standing for the backbone, N (Asn) as a polar residue, and E (Glu) as an acidic residue. The total size of the library is $3 \times 5^8 = 1.2 \times 10^6$ sequences.

Applying the PPM-based workflow to this exhaustive library, F-centered sequences are found to have the highest membrane insertion propensities, with 48.6% of all the sequences as inserters, followed by W-centered sequences, with 42.4% inserters (Fig. 5B). In contrast, Y-centered motifs show very low membrane insertion propensities, with only 1.3% inserters. This contrast between F/W-centered sequences and Y-centered sequences is consistent with the outcome of the initial set of 10 sequences.

With the large library, we can now examine the distributions of amino acids in flanking positions. Within either the membrane inserters or non-inserters, we calculated the frequencies at which each flanking position was occupied by the five types of amino acids. If the distributions of amino acids in the two subsets were totally random, the occupation frequency of any amino acid at any position would be 0.2. An occupation frequency higher than 0.2 in the inserter subset means that that amino acid promotes membrane insertion. Figure 5C displays the occupation frequencies in the inserter subset. The frequencies have the order $L > R > G > N > E$. For each amino acid, the occupation frequency is relatively uniform across all eight positions. For L and R, the frequencies in the inserter subset are > 0.2 , indicating that these amino acids promote membrane insertion. An aliphatic sidechain like L is favored in the acyl chain region, thereby stabilizing the inserted state (Fig. 3). While burying a charge in the acyl chain region is highly unfavorable, R can snorkel to form both hydrophobic interactions via its aliphatic portion and salt bridges via its terminal guanidinium group (Fig. 3A, B).

The order in occupation frequency is reversed in the non-inserter subset (Fig. 5D), where E, N, and G have values > 0.2 , meaning that they suppress membrane insertion. In contrast to R, an acidic sidechain like E is electrostatically repelled by lipid phosphates; its suppression of membrane insertion is thus expected. The suppressive effects of N and G can be attributed to the fact that it is unfavorable to position polar groups, either on the sidechain or the backbone, in the acyl chain region.

Given the much lower membrane insertion propensities of Y-centered sequences relative to F- and W-centered sequences, we report the distributions of amino acids in Fig. S8 for each aromatic center. The results for F- and W-centered sequences (Fig. S8A-D) are similar to the aggregated results shown in Fig. 5C, D. However, for Y-centered sequences, the inserter subset is highly enriched in the insertion promoter L, with corresponding depletion of insertion suppressors E, N, and G (Fig. S8E). Apparently, Y-centered sequences require substantial help from insertion promoters in flanking positions for them to become membrane inserters. In the non-inserter subset of Y-centered sequences, occupation frequencies have only very small differences from the random value of 0.2 (Fig. S8F), reflecting the fact that the non-inserter subset differs from the full set only by a small percentage (i.e., 1.3%) of sequences that belong to the inserter subset.

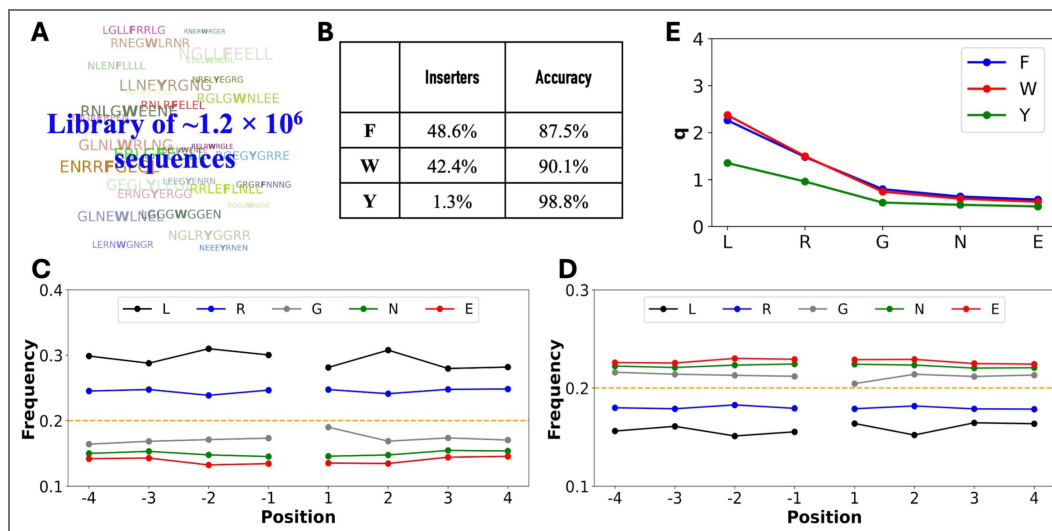


Fig. 5. Flanking residues as determinants of membrane insertion.

(A) Illustration of the library of $3 \times 5^8 = 1.2 \times 10^6$ sequences. (B) Percentages of membrane inserters and prediction accuracy of AroMIP. (C) L and R are enriched in all flanking positions of membrane inserters. (D) N, E, and G are enriched in all flanking positions of non-inserters. In (C, D), an orange dashed line is drawn at a frequency of 0.2 expected of random distributions. (E) q parameters for F-, W-, and Y-centered sequences.

The PPM results on the exhaustive library clearly demonstrate two simple rules: (1) flanking residues play key roles in determining whether an aromatic-centered motif is a membrane inserter; (2) aliphatic and basic residues promote membrane insertion, whereas polar and acidic residues suppress it. These rules motivated us to test whether a sequence-based method can predict the membrane insertion status. Based on previous models of sequence-dependent properties of IDPs, including membrane-association propensities and drug-interacting residues,^{49–52} we devised a membrane insertion score by multiplying the contributions of all flanking residues:

$$Q = \prod_{i=\pm 1}^{\pm 4} q_i$$

where q_i depends on the amino-acid type of the residue at position i with respect to the central aromatic residue. A q value > 1 corresponds to promotion of membrane insertion whereas a value < 1 corresponds to suppression of membrane insertion. We then converted the score into a membrane-insertion propensity that ranges from 0 to 1:

$$P = \frac{Q}{1 + Q} \quad (2)$$

For a given aromatic residue (F, W, or Y), we optimized the five q parameters to minimize the difference between % and the pre-defined label (0 for non-inserter and 1 for membrane inserter according to PPM) of each sequence. We call this sequence-dependent method AroMIP.

As the training size increases from 1000 to the maximum value $5^8 = 390625$, the q parameters quickly converge (Fig. S9 [↗](#)). Figure 5E [↗](#) displays the final q parameters for F-, W-, and Y-centered sequences. For all three aromatic centers, the q values follow the order $L > R > G > N > E$, as expected from the amino-acid distribution results in the inserter subsets. These values are very similar for F and W centers, > 1 for L and R but < 1 for G, N, and E. The corresponding values for the Y center are lower for all five flanking amino acids. In particular, the q value for R dips below 1, making it a weak suppressor of membrane insertion for Y-centered motifs.

As the first test of the sequence-dependent approach, we assessed how well the trained AroMIP model reproduced the PPM-based classification of the 5-amino-acid library, setting a threshold P at 0.5 for inserters (corresponding to $Q = 1$). The retrospective accuracy is 87.5%, 90.1%, and 98.8% for F-, W-, and Y-centered motifs, respectively (Fig. 5B [↗](#)). Such high accuracy levels demonstrate that AroMIP is a promising method.

AroMIP accurately predicts the membrane-insertion status of IDRs in the human proteome and identifies potential functionally important motifs

We illustrate the utility of AroMIP by applying it to all the IDRs in the human proteome.⁴⁷ From these IDRs, we collected 9-residue motifs centered on an aromatic residue, resulting in a dataset of 24308 motifs (Fig. 6A [↗](#)). We applied the PPM-based workflow to generate labels (membrane inserters or non-inserters) for all these motifs and randomly separated them into a training set (80% of motifs) and a test set (20% of motifs).

We then optimized the AroMIP model (equations [1] and [2]) on the training set, now with 20 q parameters for the full amino-acid alphabet. Figure 6B [↗](#) displays the resulting q parameters for F-, W-, and Y-centered motifs. For F- and W-centered motifs, 12 of the 20 amino acids have $q > 1$ and hence are insertion promoters at flanking positions. This group includes 3 aromatic amino acids (F, W, and Y), 6 aliphatic amino acids (L, I, M, V, P, and A), 2 basic amino acids (R and K), and a single polar amino acid (T). The q values of F and W are much higher than those of the rest amino acids, indicating unique membrane-insertion propensities. Between F- and W-centered motifs, the q values of F and W are higher in the latter; for the remaining 18 amino acids, q values are similar between F- and W-centered motifs. In comparison, all 20 q values are reduced for Y-centered

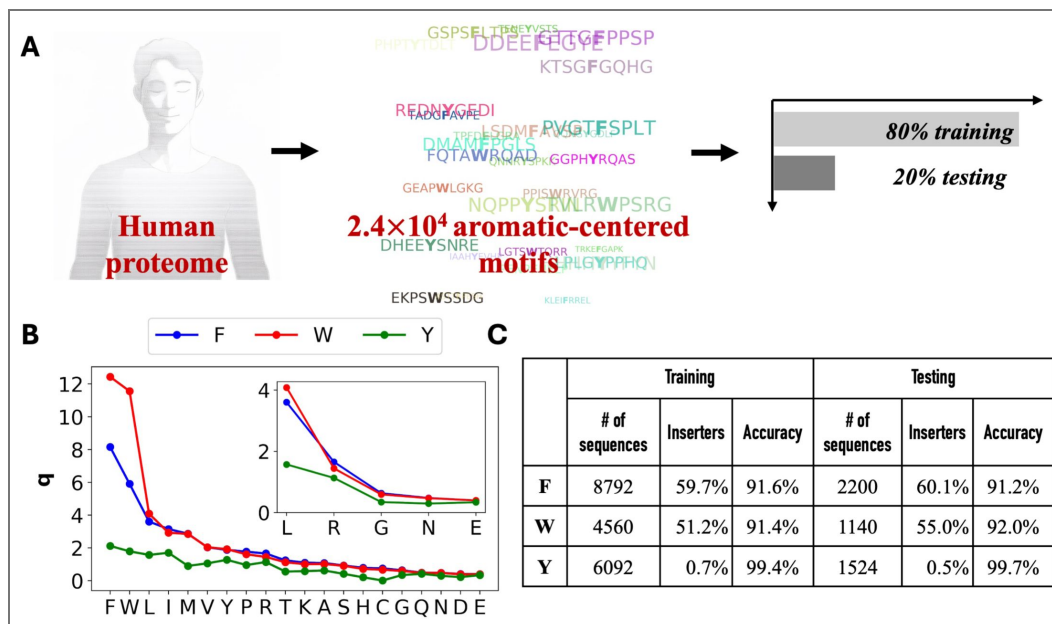


Fig 6. Implementation of AroMIP on 2.4×10^4 aromatic-centered motifs in IDRs of the human proteome.

(A) Illustration of 2.4×10^4 aromatic-centered motifs in IDRs of the human proteome. This dataset was split 80:20 for training and testing, respectively. (B) q parameters for F-, W-, and Y-centered motifs. (C) Performance of AroMIP on the training and test sets.

motifs. Consequently, 5 of the 12 amino acids are no longer in the insertion-promoting group: M, P, T, K, and A. Comparing the q values of L, R, G, N, and E here from training on a 20-amino-acid dataset (Fig. 6B inset) against those from a 5-amino-acid dataset (Fig. 5E), we see concordance in several respects, including the same rank order $L > R > G > N > E$, similarity between F- and W-centered motifs and contrast with Y-centered motifs, and same groupings of insertion promoters and suppressors, except for R in Y-centered motifs, which switches from a weak suppressor in the 5-amino-acid dataset to a weak promoter in the 20-amino-acid dataset. Quantitatively, for these five amino acids, the range of q values obtained from the 20-amino-acid dataset is widened somewhat, at both the higher (for L) and lower (for E) ends.

Now let us look at the performance of the trained AroMIP model. In the training set, the percentages of membrane inserters are 59.7%, 51.2%, and 0.7% for F-, W-, and Y-centered motifs, respectively (Fig. 6C). The retrospective accuracy is 91.6%, 91.4%, and 99.4%, respectively. In the test set, the percentages of membrane inserters are 60.1%, 55.0%, and 0.5% for F-, W-, and Y-centered motifs, respectively, which are, as expected, comparable to the counterparts in the training set. The prospective accuracy is 91.2%, 92.0%, and 99.7%, respectively. The latter are essentially identical to the counterparts on the training set, indicating the robustness of the AroMIP model.

We list the top-scoring aromatic-centered motifs in Supplementary Table S1. The F- and W-centered motifs reach the full value of 1 in membrane-insertion propensity, and are typically flanked by two other F and W residues and a combination of aliphatic and basic residues. In contrast, the Y-centered motifs are limited to 0.78 in P, and are flanked by one F or W but a high number of aliphatic residues. Some of these motifs may actually be functionally important. In particular, the transmembrane protein anoctamin-8 resides in the endoplasmic reticulum and tethers the plasma membrane via its C-terminal IDR to assemble Ca^{2+} -signaling complexes.⁵³ The $\text{A}_{1023}\text{FLSFKFLK}_{1031}$ motif in this IDR may insert into the plasma membrane to facilitate this function. Similarly, GARRE1 localizes to the plasma membrane and intracellular vesicles, likely via a truncated BAR domain, to mediate ciliogenesis.⁵⁴ Membrane insertion by $\text{R}_{885}\text{TWPFPEFF}_{893}$ in the C-terminal IDR may fine-tune this function. Sec16B, with a key role in biogenesis of peroxisomes, tightly associates with endoplasmic reticulum membranes via a central conserved domain but with help from its C-terminal IDR.⁵⁵ The latter may involve membrane insertion by the $\text{S}_{910}\text{GFGWFSWF}_{918}$ motif.

Additional test cases provide further validation of AroMIP

We have implemented AroMIP as a web server, accessible at <https://zhougroup-uic.github.io/AroMIP/>, and now report results on additional test cases. First, we go back to our initial set of 10 sequences. The Q and P values for these sequences are listed in Table 1, which result in correct predictions of sequences 1 and 2 as membrane inserters and the other 8 as non-inserters. Moreover, the raw P values are highly correlated with both the percentages of membrane-inserting MD simulations (Fig. S10A) and the percentages of membrane-inserting conformations in PPM runs (Fig. S10B). This correlation suggests that AroMIP can be used not only for binary classification but also for detecting fine differences between motifs.

Next we report the performance of AroMIP on aromatic-centered motifs in 12 other IDPs or IDRs for which information regarding membrane insertion is available (Supplementary Table S2). Using solid-state NMR, Zhang et al.¹⁰ observed strong cross peaks between acyl chains and five phenol rings from a 25-residue IDR (residues 152-176) of (MARCKS). AroMIP predicts very high membrane-insertion propensities for four F residues and high P for the fifth F residue, consistent with the solid-state NMR results. Similarly, solution NMR found cross peaks between the indole ring of W11 in the 86-residue IDP Tat and acyl chains.⁴¹ AroMIP correctly predicts W11 as a membrane inserter with a high P (= 0.73). In comparison, four other aromatic residues are predicted with lower P values, with one (F38) passing the 0.5 threshold.

Attenuation of NMR signals indicated strong membrane binding by Y511 and weaker binding by Y483 in the C-terminal IDR of Tir.¹⁷ Whether these Y sidechains were inserted into the acyl chain region was unclear. AroMIP predicts both residues as non-inserters, but with a higher membrane-

insertion propensity for Y511 ($P = 0.14$) than for Y483 ($P = 0.01$). A slightly higher propensity is predicted for F489 ($P = 0.35$). Interestingly, these P values correlate with the extents of NMR signal attenuation upon binding DMPG:DHPC membranes. A similar correlation is found for three aromatic residues in the cytoplasmic IDR of mGluR3, with high NMR signal attenuation¹⁶ and high membrane-insertion propensity ($P = 0.72$) for F844 and low signal attenuation and low P values (i.e., 0.002 and 0.04) for Y853 and Y861. Moreover, the Y853F mutation increases both the extent of NMR signal attenuation and the membrane-insertion propensity (from 0.002 to 0.33). AroMIP predicts Y853 as a non-inserter, which is consistent with all-atom MD simulations showing that this residue only makes excursions into the acyl chain region.¹⁶

All-atom MD simulations showed that the aromatic motif $\mathbf{W}_{440}\mathbf{FFGW}_{444}$ in the 57-residue IDR of Bap1 is stably inserted into the acyl chain region.²⁰ AroMIP predicts extremely high propensities (close to 1) for these four aromatic residues, due to the presence of flanking F or W (see Fig. 6B [↗](#)). Three other W residues (at positions 420, 432, and 456) are also predicted as inserters (P values range from 0.78 to 0.86); MD simulations did not find membrane insertion by these residues. For KtrB, all-atom simulations showed membrane insertion by F10 in the N-terminal IDR (residues 1-24), but not the adjacent Y11.⁹ In complete agreement, AroMIP predicts a high membrane-insertion propensity ($P = 0.75$) for F10 but a very low P value (0.07) for Y11. All-atom simulations also found membrane insertion by W21, W25, and W29 in the N-terminal IDR (residues 1-33) of ABHD5, with a greater depth for W25 than for W21 and W29.⁴⁴ AroMIP predicts a perfect P value of 1 for W25, due to the presence of the other two W residues in flanking positions, and somewhat lower values for W21 and W29. Coarse-grained simulations showed membrane insertion by the $\mathbf{I}_{267}\mathbf{FPPV}_{271}$ motif in the membrane-proximal IDR (residues 260-280) of the prolactin receptor;¹⁴ AroMIP predicts a very high membrane-insertion propensity ($P = 0.99$) for F268.

Cryo-EM structures of OPA1 place W775 and W778 in the membrane-inserting loop deeply into the acyl chain region of cardiolipin-containing membranes.^{35, 36} Likewise, cryo-EM structures suggest that aromatic residues at the tip of the $\beta 1$ - $\beta 2$ loop in gasdermin D ($\mathbf{W}_{48}\mathbf{FW}_{50}$) and gasdermin B ($\mathbf{F}_{46}\mathbf{F}_{47}$) insert into the acyl chain region as an anchor.^{37, 39} This scenario was demonstrated in all-atom MD simulations.³⁸ AroMIP predicts very high membrane-insertion propensities for all these aromatic residues. In contrast, a medium propensity ($P = 0.55$) is predicted for W41 in a loop of PLAT1, consistent with its transient insertion in all-atom MD simulations.⁴⁰

Discussion

We have used three complementary approaches to decipher the membrane insertion code of aromatic-centered motifs in IDPs and IDRs. All-atom MD simulations revealed the characteristics of membrane-bound states and the pathways leading to membrane insertion. F- and W-centered motifs stably insert into the acyl chain region, but a Y-centered motif does so only transiently. Flanking aliphatic sidechains make major contributions to membrane binding and insertion of the motif by penetrating into the acyl chain region; flanking basic sidechains draw the motif to the membrane surface via long-range electrostatic interactions, keep it at the surface via salt bridges, and stabilize the inserted state via salt bridges and snorkeling. The PPM-based workflow enabled a full exploration of sequence space and the discovery that, for membrane insertion by F- and W-centered motifs, flanking F and W are strong promoters, aliphatic sidechains are medium promoters, and R is a modest promoter. Moreover, Y-centered motifs have very low membrane-insertion propensities. The sequence-based AroMIP method accurately captures these rules and enables fast prediction of membrane-insertion status.

The 20 q parameters (Fig. 6B [↗](#)) embody the membrane insertion code. For F/W-centered motifs, flanking F and W have very high q values, highlighting their unique ability in membrane insertion and again justifying our focus on aromatic-centered motifs. Consequently, F/W clusters result in very high membrane-insertion propensities. This pattern is seen in the top-scoring F/W-centered motifs from IDRs of the human proteome (Supplementary Table S1 [↗](#)), in the $\mathbf{W}_{440}\mathbf{FFGW}_{444}$ motif of Bap1, the $\mathbf{W}_{21}\mathbf{LTGWLPTW}_{29}$ motif of ABHD5, and the $\mathbf{W}_{48}\mathbf{FW}_{50}$ motif of gasdermin D. For Bap1, mutating WFFG to LGPE significantly reduced membrane binding.²⁰ Similarly, mutating W48 and W50 to E (or G in the mouse homolog) rendered gasdermin D defective in pore formation.^{37, 56} It

will be interesting to see whether the predicted effects of F/W clusters in the C-terminal IDRs of anoctamin-8, GARRE1, and Sec16B pan out. Whereas F/W clustering can ensure membrane insertion, aliphatic and basic residues can tune the membrane-insertion propensity to desired values. Aliphatic and basic residues can regulate membrane insertion in distinct ways. While aliphatic residues have a somewhat higher ability to stabilize the inserted state, basic residues can draw the aromatic-centered motif to the membrane surface prior to insertion.

The most unexpected result is the enormous difference in membrane-insertion propensity between Y and F or W. In both the exhaustive 5-amino-acid library and the dataset from IDRs of the human proteome, F/W-centered motifs have ~50% chance of being a membrane inserter, but Y-centered motifs have only ~1% chance of being an inserter (Figs. 5B and 6C). Burial of the tip hydroxyl of Y by itself in the acyl chain region is highly unfavorable; instead, the hydroxyl hydrogen bonds with a glycerol oxygen atom, and an excursion into the acyl chain region is accompanied by the glycerol group, creating local curvature in the membrane surface (Fig. 3C). Although the W sidechain also has a polar NH group, its position in the five-membered ring allows it to hydrogen bond with a glycerol oxygen atom while the six-membered ring is buried in the acyl chain region (Fig. 2B). Interestingly, the low membrane-insertion propensity of Y fits well with its roles in signal transduction, where a Y-centered motif has to switch binding partners between a membrane and an effector protein.^{11, 15–17} Stable deep insertion would hinder this switch. Indeed, in many cases, the switch from membrane to effector protein requires phosphorylation of the central Y residue, which weakens membrane binding. Deep insertion would also hamper phosphorylation by limiting the kinase's access to the Y residue.

We have thoroughly validated AroMIP. The method was motivated by distinct distributions of aromatic, aliphatic, basic, polar, and acidic residues in flanking positions of membrane-inserting aromatic-centered motifs, and quantitatively accounts for their disparate contributions to the membrane-insertion propensity of the central aromatic residue. When tested on IDRs in the human proteome, the accuracy of AroMIP exceeds 90% for F/W-centered motifs and approaches 100% for Y-centered motifs. Additional test cases, previously characterized by experimental or MD studies, provide further validation. We also predict possible functionally important motifs in the IDRs of anoctamin-8, GARRE1, and Sec16B. While these predictions are yet to be tested, they illustrate AroMIP's value in narrowing the search for key residues in functional studies. AroMIP can likewise guide the design of MD simulation studies, in particular, by focusing attention on putative membrane-inserting motifs. With additional applications of AroMIP and further development, we anticipate that the understanding of membrane insertion will reach the same clarity as for the other modes (via amphipathic helices and polybasic motifs) of IDP-membrane association.

Methods

MD simulations

Each of the ten nine-residue peptides from the initial dataset in a disordered conformation generated through TraDES⁴⁸ was placed near a membrane, with the central aromatic residue just above the phosphate plane. The peptides were N-terminally capped with an acetyl group and C-terminally capped with an amide group to mimic the state of the motifs within the full-length protein. The upper leaflet of the membrane was composed of 70 POPC, 25 POPS, and 5 PIP₂ lipids; the lower leaflet was composed of 97-99 POPC lipids, with matching surface areas between the two leaflets. The system was then solvated in a rectangular box (with dimensions listed in Supplementary Table S3) with water and NaCl (neutralizing ions plus 150 mM) using the CHARMM-GUI web server.⁵⁷ The force fields were CHARMM36⁵⁸ for proteins and TIP3P⁵⁹ for water. After energy minimization for 10000 steps, each system was equilibrated in six steps in NAMD 3.0.⁶⁰ The first two were at constant temperature and volume, while the remaining four were at constant temperature and pressure. The timesteps were 1 fs in the first three and increased to 2 fs in the last three steps. The simulation times were 125 ps in the first three and 500 ps for the last three steps. During these six steps, restraints on the lipid phosphorus atoms were

gradually reduced with the force constant changing from 5 kcal/mol/Å² to 0; similarly, the force constant on peptide heavy atoms was reduced from 10 kcal/mol/Å² to 0. Production runs, with 20 replicates for each peptide, were carried out at constant temperature and pressure without restraints and a 2 fs timestep using pmemd.cuda⁶¹ in AMBER 22⁶², for a simulation time of 1 μs. Bond lengths containing hydrogen were constrained using the SHAKE algorithm.⁶³ Long-range electrostatic interactions were treated using the particle mesh Ewald method⁶⁴ with a nonbonded cutoff of 12 Å. The temperature was maintained at 300 K by the Langevin thermostat⁶⁵ with a damping constant of 1 ps⁻¹; the pressure was maintained at 1 bar by the Berendsen barostat.⁶⁶ Snapshots were saved every 100 ps for analysis.

MD and PPM data analysis

Z_{tip} distances were calculated for the last 100 ns of simulations using a custom Python code by measuring the difference between the Z coordinate of the tip carbon atom of an aromatic sidechain and the average Z coordinate of phosphorous atoms in the upper leaflet. Frames in which the peptide moved to the image membrane (Z_{tip} beyond the midpoint between the upper leaflet and its periodic image) were excluded from analysis. Distributions were calculated using only the remaining frames. The tilt angle of the aromatic sidechain was the polar angle of a vector connecting two diagonal atoms on the six-membered ring (C γ -C ζ in F and Y and C δ_2 -C η_2 in W; Fig. S2A [↗](#)). A peptide was considered membrane-interacting when any of its heavy atom was within 5 Å of any heavy atom of the membrane. Cation- π interaction between a lipid headgroup and an aromatic sidechain was defined when the distances between a choline nitrogen and carbon atoms on the six-membered ring in the aromatic sidechain were less than 7 Å and these distances differed by less than 1.5 Å (<https://github.com/weberdak/flexible-interaction-tool/blob/master/flirt.tcl> [↗](#); code written by D. K. Weber).⁶⁷ Ca chemical shifts were calculated every 1 ns over the last 100 ns of simulations using SHIFTX2.⁶⁸ After averaging the 100 frames in a simulation and then over three selected simulations, the corresponding random-coil values were subtracted to yield Ca secondary chemical shifts.

PPM runs used the source code from https://console.cloud.google.com/storage/browser/opm-assets/ppm3_code [↗](#),⁴⁶ with a batch of 1000 or 100 conformations generated by TraDES as input for each peptide. The output for each peptide conformation was the membrane interaction free energy (ΔG) and a coordinate file representing the peptide's pose relative to the membrane. Here Z_{tip} was defined using the most buried ring atom on the central aromatic residue.

Aromatic-centered dataset from IDRs of the human proteome

All IDR sequences in the human proteome were collected from <https://mobidb.org/> [↗](#).⁴⁷ Each IDR sequence was then scanned for 9-residue motifs centered at F, W, or Y; duplicates were removed, yielding a dataset of 24308 sequences.

AroMIP implementation

For each peptide sequence, we calculated a membrane insertion score by multiplying the contributions of all flanking residues (see equation [1]). We then converted the score to a membrane insertion propensity ranging from 0 to 1 (see equation [2]). For a given aromatic residue (F, W, or Y), the q parameters were optimized by using the `scipy.optimize` module.⁶⁹ in Python to minimize the difference between the predicted inserted propensity and the pre-defined label (0 for non-inserter and 1 for membrane inserter) for sequences in a training set. The resulting AroMIP model was then applied to a test set for reporting accuracy and implemented as a web server. A threshold propensity of 0.5 separated membrane inserters from non-inserters.

Data availability

All source data are deposited on Github: <https://github.com/hzhou43/AroMIP> [↗](#).

Acknowledgements

This work was supported by National Institutes of Health Grant GM118091.

Additional files

[Supplementary Tables and Figures](#) 

Additional information

Funding

Funder	Grant reference number	Author
HHS NIH National Institute of General Medical Sciences (NIGMS)	GM118091	Huan-Xiang Zhou

Author ORCID iDs

Huan-Xiang Zhou:  <https://orcid.org/0000-0001-9020-0302>

References

- 1 **Busch DJ**, Houser JR, Hayden CC, Sherman MB, Lafer EM, Stachowiak JC (2015) Intrinsically disordered proteins drive membrane curvature. *Nat Commun* **6**:7875 <https://doi.org/10.1038/ncomms8875> | [PubMed](#)
- 2 **Snead WT**, Hayden CC, Gadok AK, Zhao C, Lafer EM, Rangamani P, Stachowiak JC (2017) Membrane fission by protein crowding. *Proc Natl Acad Sci U S A* **114**:E3258 <https://doi.org/10.1073/pnas.1616199114> | [PubMed](#)
- 3 **Zeno WF**, Baul U, Snead WT, DeGroot ACM, Wang L, Lafer EM, Thirumalai D, Stachowiak JC (2018) Synergy between intrinsically disordered domains and structured proteins amplifies membrane curvature sensing. *Nat Commun* **9**:4152 <https://doi.org/10.1038/s41467-018-06532-3> | [PubMed](#)
- 4 **Zeno WF**, Thatte AS, Wang L, Snead WT, Lafer EM, Stachowiak JC (2019) Molecular Mechanisms of Membrane Curvature Sensing by a Disordered Protein. *J Am Chem Soc* **141**:10361 <https://doi.org/10.1021/jacs.9b03927> | [PubMed](#)
- 5 **Mahajan M**, Bharambe N, Shang Y, Lu B, Mandal A, Madan Mohan P, Wang R, Boatz JC, Galvez J Manuel Martinez, Shnyrova AV, *et al.* (2021) NMR identification of a conserved Drp1 cardiolipin-binding motif essential for stress-induced mitochondrial fission. *Proc Natl Acad Sci U S A* **118** <https://doi.org/10.1073/pnas.2023079118> | [PubMed](#)
- 6 **Peng R**, Rochon K, Hutson AN, Stagg SM, Mears JA (2025) The Structure of the Drp1 Lattice on Membrane. *J Mol Biol* **437**:169125 <https://doi.org/10.1016/j.jmb.2025.169125> | [PubMed](#)
- 7 **Qin S**, Hicks A, Dey S, Prasad R, Zhou HX (2022) ReSMAP: Web Server for Predicting Residue-Specific Membrane-Association Propensities of Intrinsically Disordered Proteins. *Membranes* **12**:773 <https://doi.org/10.3390/membranes12080773> | [PubMed](#)
- 8 **Goretzki B**, Wiedemann C, McCray BA, Schäfer SL, Jansen J, Tebbe F, Mitrovic S.-A, Nöth J, Cabezudo AC, Donohue JK, *et al.* (2023) Crosstalk between regulatory elements in disordered TRPV4 N-terminus modulates lipid-dependent channel activity. *Nat Commun* **14**:4165 <https://doi.org/10.1038/s41467-023-39808-4> | [PubMed](#)
- 9 **Stautz J**, Griwatz D, Kaltwasser S, Mehdipour AR, Ketter S, Thiel C, Wunnicke D, Schrecker M, Mills DJ, Hummer G, *et al.* (2025) A short intrinsically disordered region at KtrB's N-terminus facilitates allosteric regulation of K^+ channel KtrAB. *Nat Commun* **16**:4252 <https://doi.org/10.1038/s41467-025-59546-z> | [PubMed](#)

- 10 Zhang W, Crocker E, McLaughlin S, Smith SO (2003) Etk/Bmx transactivates vascular endothelial growth factor 2 and recruits phosphatidylinositol 3-kinase to mediate the tumor necrosis factor-induced angiogenic pathway. *J Biol Chem* **278**:21459 <https://doi.org/10.1074/jbc.m310678200> | PubMed
- 11 Sigalov AB, Hendricks GM (2009) Membrane binding mode of intrinsically disordered cytoplasmic domains of T cell receptor signaling subunits depends on lipid composition. *Biochem Biophys Res Commun* **389**:388 <https://doi.org/10.1016/j.bbrc.2009.09.002> | PubMed
- 12 Yan X, Walkiewicz M, Carlson J, Leiphon L, Grove B (2009) Gravin dynamics regulates the subcellular distribution of PKA. *Exp Cell Res* **315**:1247 <https://doi.org/10.1016/j.yexcr.2008.12.026> | PubMed
- 13 Wang Z, Fan H, Hu X, Khamo J, Diao J, Zhang K, Pogorelov TV (2019) Coaction of Electrostatic and Hydrophobic Interactions: Dynamic Constraints on Disordered TrkA Juxtamembrane Domain. *J Phys Chem B* **123**:10709 <https://doi.org/10.1021/acs.jpcc.9b09352> | PubMed
- 14 Araya-Secchi R, Bugge K, Seiffert P, Petry A, Haxholm GW, Lindorff-Larsen K, Pedersen SF, Arleth L, Kragelund BB (2023) The prolactin receptor scaffolds Janus kinase 2 via co-structure formation with phosphoinositide-4,5-bisphosphate. *eLife* **12**:e84645 <https://doi.org/10.7554/eLife.84645> | PubMed
- 15 Jin H, Liu D, Ni Y, Wang H, Long D (2024) Luminescent Fe(III) Complex Sensitizes Aerobic Photon Upconversion and Initiates Photocatalytic Radical Polymerization. *J Am Chem Soc* **146**:791 <https://doi.org/10.1021/jacs.4c14248> | PubMed
- 16 Mancinelli CD, Marx DC, Gonzalez-Hernandez AJ, Huynh K, Mancinelli L, Arefin A, Khelashvili G, Levitz J, Eliezer D (2024) Control of G protein-coupled receptor function via membrane-interacting intrinsically disordered C-terminal domains. *Proc Natl Acad Sci U S A* **121**:e2407744121 <https://doi.org/10.1073/pnas.2407744121> | PubMed
- 17 Vieira MFM, Hernandez G, Zhong Q, Arbesu M, Veloso T, Gomes T, Martins ML, Monteiro H, Frazao C, Frankel G, *et al.* (2024) The pathogen-encoded signalling receptor Tir exploits host-like intrinsic disorder for infection. *Commun Biol* **7**:179 <https://doi.org/10.1038/s42003-024-05856-9> | PubMed
- 18 Dey S, Zhou HX (2022) Membrane Tethering of SepF, a Membrane Anchor for the Mycobacterium tuberculosis Z-ring. *J Mol Biol* **434**:167817 <https://doi.org/10.1016/j.jmb.2022.167817> | PubMed
- 19 Smrt ST, Escobar CA, Dey S, Cross TA, Zhou HX (2023) An Arg/Ala-rich helix in the N-terminal region of M. tuberculosis FtsQ is a potential membrane anchor of the Z-ring. *Commun Biol* **6**:311 <https://doi.org/10.1038/s42003-023-04686-5> | PubMed
- 20 Huang X, Prasad R, Saluja S, Yang Y, Yan Q, Shuster SO, Karatekin E, Olson R, Lin C, Davis CM, *et al.* (2026) Correction: Structural basis for substrate recognition and inhibition of thioredoxin glutathione reductase from *Schistosoma japonicum*: Implications for antiparasitic development. *PLoS Pathog* **22**:e1013990 <https://doi.org/10.1371/journal.ppat.1014386> | PubMed
- 21 Rombach J, Nielsen TTE, Schmidt JH, Agenant J, Larsen AH, Madsen KL (2025) A proteome-wide screen for membrane-interactions in intrinsically disordered regions of transmembrane proteins reveals a role in disease. *bioRxiv* <https://doi.org/10.1101/2025.08.15.670465>
- 22 MacAinsh M, Muhammedkutty FNK, Prasad R, Zhou HX (2025) Membrane Association of Intrinsically Disordered Proteins. *Annu Rev Biophys* **54**:275 <https://doi.org/10.1146/annurev-biophys-070124-092816> | PubMed
- 23 Sapay N, Guermeur Y, Deleage G (2006) Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics* **7**:255 <https://doi.org/10.1186/1471-2105-7-255> | PubMed
- 24 Wimley WC, White SH (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol* **3**:842 <https://doi.org/10.1038/nsb1096-842> | PubMed
- 25 Ulmschneider MB, Sansom MS (2001) Amino acid distributions in integral membrane protein structures. *Biochim Biophys Acta* **1512**:1 [https://doi.org/10.1016/s0005-2736\(01\)00299-1](https://doi.org/10.1016/s0005-2736(01)00299-1) | PubMed

- 26 **Hong H**, Park S, Jiménez RH, Rinehart D, Tamm LK (2007) Copper(I)-alpha-ketocarboxylate complexes: characterization and O₂ reactions that yield copper-oxygen intermediates capable of hydroxylating arenes. *J Am Chem Soc* **129**:8320 <https://doi.org/10.1021/ja0760426> | PubMed
- 27 **McDonald SK**, Fleming KG (2016) Aromatic Side Chain Water-to-Lipid Transfer Free Energies Show a Depth Dependence across the Membrane Normal. *J Am Chem Soc* **138**:7946 <https://doi.org/10.1021/jacs.6b03460> | PubMed
- 28 **Gaede HC**, Yau WM, Gawrisch K (2005) Electrostatic contributions to indole-lipid interactions. *J Phys Chem B* **109**:13014 <https://doi.org/10.1021/jp0511000> | PubMed
- 29 **Norman KE**, Nymeyer H (2006) Indole localization in lipid membranes revealed by molecular simulation. *Biophys J* **91**:2046 <https://doi.org/10.1529/biophysj.105.080275> | PubMed
- 30 **MacCallum JL**, Bennett WF, Tieleman DP (2008) Distribution of amino acids in a lipid bilayer from computer simulations. *Biophys J* **94**:3393 <https://doi.org/10.1529/biophysj.107.112805> | PubMed
- 31 **Blaser G**, Sanderson JM, Wilson MR (2009) Free-energy relationships for the interactions of tryptophan with phosphocholines. *Org Biomol Chem* **7**:5119 <https://doi.org/10.1039/b913919b> | PubMed
- 32 **Waheed Q**, Khan HM, He T, Roberts M, Gershenson A, Reuter N (2019) Interfacial Aromatics Mediating Cation- π Interactions with Choline-Containing Lipids Can Contribute as Much to Peripheral Protein Affinity for Membranes as Aromatics Inserted below the Phosphates. *J Phys Chem Lett* **10**:3972 <https://doi.org/10.1021/acs.jpcllett.9b01639> | PubMed
- 33 **Tubiana T**, Sillitoe I, Orengo C, Reuter N (2022) Dissecting peripheral protein-membrane interfaces. *PLoS Comput Biol* **18**:e1010346 <https://doi.org/10.1371/journal.pcbi.1010346> | PubMed
- 34 **He T**, Gershenson A, Eyles SJ, Lee YJ, Liu WR, Wang J, Gao J, Roberts MF (2015) Fluorinated Aromatic Amino Acids Distinguish Cation- π Interactions from Membrane Insertion. *J Biol Chem* **290**:19334 <https://doi.org/10.1074/jbc.M115.668343> | PubMed
- 35 **Nyenhuis SB**, Wu X, Strub MP, Yim YI, Stanton AE, Baena V, Syed ZA, Canagarajah B, Hammer JA, Hinshaw JE (2023) OPA1 helical structures give perspective to mitochondrial dysfunction. *Nature* **620**:1109 <https://doi.org/10.1038/s41586-023-06462-1> | PubMed
- 36 **von der Malsburg A**, Sapp GM, Zuccaro KE, von Appen A, Moss FR, Kalia R, Bennett JA, Abriata LA, Dal Peraro M, van der Laan M, *et al.* (2023) Structural mechanism of mitochondrial membrane remodelling by human OPA1. *Nature* **620**:1101 <https://doi.org/10.1038/s41586-023-06441-6> | PubMed
- 37 **Xia S**, Zhang Z, Magupalli VG, Pablo JL, Dong Y, Vora SM, Wang L, Fu TM, Jacobson MP, Greka A, *et al.* (2021) Gasdermin D pore structure reveals preferential release of mature interleukin-1. *Nature* **593**:607 <https://doi.org/10.1038/s41586-021-03478-3> | PubMed
- 38 **Schaefer SL**, Hummer G (2022) Sublytic gasdermin-D pores captured in atomistic molecular simulations. *eLife* **11** <https://doi.org/10.7554/elife.81432> | PubMed
- 39 **Wang C**, Shivcharan S, Tian T, Wright S, Ma D, Chang J, Li K, Song K, Xu C, Rathinam VA, *et al.* (2023) Deforming lanthanum trihydride for superionic conduction. *Nature* **616**:590 <https://doi.org/10.1038/s41586-023-05815-0> | PubMed
- 40 **Kulke M**, Kurtz E, Boren DM, Olson DM, Koenig AM, Hoffmann-Benning S, Vermaas JV (2024) PLAT domain protein 1 (PLAT1/PLAFP) binds to the Arabidopsis thaliana plasma membrane and inserts a lipid. *Plant Sci* **338**:111900 <https://doi.org/10.1016/j.plantsci.2023.111900> | PubMed
- 41 **Ghanam RH**, Eastep GN, Saad JS (2023) Structural Insights into the Mechanism of HIV-1 Tat Secretion from the Plasma Membrane. *J Mol Biol* **435**:167880 <https://doi.org/10.1016/j.jmb.2022.167880> | PubMed
- 42 **Xu C**, Gagnon E, Call ME, Schnell JR, Schwieters CD, Carman CV, Chou JJ, Wucherpfennig KW (2008) Regulation of T cell receptor activation by dynamic membrane binding of the CD3epsilon cytoplasmic tyrosine-based motif. *Cell* **135**:702 <https://doi.org/10.1016/j.cell.2008.09.044> | PubMed

- 43 Prakaash D, Cook GP, Acuto O, Kalli AC (2021) Multi-scale simulations of the T cell receptor reveal its lipid interactions, dynamics and the arrangement of its cytoplasmic region. *PLoS Comput Biol* **17**:e1009232 <https://doi.org/10.1371/journal.pcbi.1009232> | PubMed
- 44 Kumar A, Sanders M, Zhang H, Zhou L, Parveen S, Kelly CV, Granneman JG, Huang YM (2025) Unraveling the Molecular Mechanisms of ABHD5 Membrane Targeting. *bioRxiv* <https://doi.org/10.1101/2025.05.06.652485>
- 45 Muhammedkutty FNK, MacAinsh M, Zhou HX (2025) Atomistic molecular dynamics simulations of intrinsically disordered proteins. *Curr Opin Struct Biol* **92**:103029 <https://doi.org/10.1016/j.sbi.2025.103029> | PubMed
- 46 Lomize AL, Pogozheva ID, Lomize MA, Mosberg HI (2006) Positioning of proteins in membranes: a computational approach. *Protein Sci* **15**:1318 <https://doi.org/10.1110/ps.062126106> | PubMed
- 47 Piovesan D, Del Conte A, Clementel D, Monzon AM, Bevilacqua M, Aspromonte MC, Iserte JA, Orti FE, Marino-Buslje C, Tosatto SCE (2023) MobiDB: 10 years of intrinsically disordered proteins. *Nucleic Acids Res* **51**:D438 <https://doi.org/10.1093/nar/gkac1065> | PubMed
- 48 Feldman HJ, Hogue CW (2000) A fast method to sample real protein conformational space. *Proteins* **39**:112 [https://doi.org/10.1002/\(sici\)1097-0134\(20000501\)39:2<112::aid-prot2>3.0.co;2-b](https://doi.org/10.1002/(sici)1097-0134(20000501)39:2<112::aid-prot2>3.0.co;2-b) | PubMed
- 49 Li D.-W, Xie M, Brüsweiler R (2020) Intrinsic Activity of Metal Centers in Metal-Nitrogen-Carbon Single-Atom Catalysts for Hydrogen Peroxide Synthesis. *J Am Chem Soc* **142**:10730 <https://doi.org/10.1021/jacs.0c10636> | PubMed
- 50 Qin S, Hicks A, Dey S, Prasad R, Zhou HX (2022) ReSMAP: Web Server for Predicting Residue-Specific Membrane-Association Propensities of Intrinsically Disordered Proteins. *Membranes* **12** <https://doi.org/10.3390/membranes12080773> | PubMed
- 51 Qin S, Zhou HX (2024) Pharyngeal mechanosensory neurons control food swallow in *Drosophila melanogaster*. *eLife* **12** <https://doi.org/10.7554/elife.88614> | PubMed
- 52 MacAinsh M, Qin S, Zhou HX (2025) DIRseq as a method for predicting drug-interacting residues of intrinsically disordered proteins from sequences. *eLife* **14** <https://doi.org/10.7554/elife.107470> | PubMed
- 53 Jha A, Chung WY, Vachel L, Maleth J, Lake S, Zhang G, Ahuja M, Muallem S (2019) Anoctamin 8 tethers endoplasmic reticulum and plasma membrane for assembly of Ca²⁺ signaling complexes at the ER/PM compartment. *EMBO J* **38** <https://doi.org/10.15252/emboj.2018101452> | PubMed
- 54 Serres MP, Shaughnessy R, Escot S, Hammich H, Cuvelier F, Salles A, Rocancourt M, Verdon Q, Gaffuri AL, Sourigues Y, *et al.* (2023) MiniBAR/GARRE1 is a dual Rac and Rab effector required for ciliogenesis. *Dev Cell* **58**:2477 <https://doi.org/10.1016/j.devcel.2023.09.010> | PubMed
- 55 Yonekawa S, Furuno A, Baba T, Fujiki Y, Ogasawara Y, Yamamoto A, Tagaya M, Tani K (2011) Sec16B is involved in the endoplasmic reticulum export of the peroxisomal membrane biogenesis factor peroxin 16 (Pex16) in mammalian cells. *Proc Natl Acad Sci U S A* **108**:12746 <https://doi.org/10.1073/pnas.1103283108> | PubMed
- 56 Liu Z, Wang C, Yang J, Zhou B, Yang R, Ramachandran R, Abbott DW, Xiao TS (2019) Targeting Degradation of the Transcription Factor C/EBP β Reduces Lung Fibrosis by Restoring Activity of the Ubiquitin-Editing Enzyme A20 in Macrophages. *Immunity* **51**:43 <https://doi.org/10.1016/j.immuni.2019.06.014> | PubMed
- 57 Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**:1859 <https://doi.org/10.1002/jcc.20945> | PubMed
- 58 Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell AD, Pastor RW (2010) Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* **114**:7830 <https://doi.org/10.1021/jp101759q> | PubMed
- 59 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**:926 <https://doi.org/10.1063/1.445869>

- 60 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**:1781 <https://doi.org/10.1002/jcc.20289> | PubMed
- 61 Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **9**:3878 <https://doi.org/10.1021/ct400314y> | PubMed
- 62 Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, *et al.* (2022) Amber22.
- 63 Ryckaert J.-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **23**:327 [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
- 64 Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* **103**:8577 <https://doi.org/10.1063/1.470117>
- 65 Pastor RW, Brooks BR, Szabo A (1988) An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol Phys* **65**:1409 <https://doi.org/10.1080/00268978800101881>
- 66 Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**:3684 <https://doi.org/10.1063/1.448118>
- 67 Grauffel C, Yang B, He T, Roberts MF, Gershenson A, Reuter N (2013) Cation- π interactions as lipid-specific anchors for phosphatidylinositol-specific phospholipase C. *J Am Chem Soc* **135**:5740 <https://doi.org/10.1021/ja312656v> | PubMed
- 68 Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* **50**:43 <https://doi.org/10.1007/s10858-011-9478-4> | PubMed
- 69 Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, *et al.* (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**:261 <https://doi.org/10.1038/s41592-019-0686-2> | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

This work investigates the membrane insertion of aromatic-centered sequences in IDPs. Using a combination of all-atom MD simulations, the PPM method, and development of the sequence-based predictor AroMIP, the authors aim to establish a quantitative membrane insertion role for aromatic-centered motifs. The study demonstrates that flanking aliphatic and basic residues promote membrane insertion, whereas acidic and polar residues suppress insertion, and further reveals a difference between F/W-centered motifs and Y-centered motifs. The resulting AroMIP model achieves high predictive accuracy on human IDPs and is implemented as a publicly accessible web server.

Strengths:

This work addresses an important biological problem, as aromatic-driven membrane insertion remains poorly characterized despite mediating diverse functions like membrane remodeling and signaling. A key strength is the combination of complementary approaches, e.g., MD simulations provide mechanistic insight into insertion pathways, while PPM enables exhaustive sequence space exploration. The large-scale analysis clearly establishes L and R as promoters and E, N, and G as suppressors. The work also provides valuable mechanistic insight into how aromatic, aliphatic, and basic residues cooperate to stabilize membrane insertion states. Another important strength is the development of AroMIP as a practical prediction tool with a user-friendly online server that appears computationally efficient and

broadly accessible to the community. The work is also well connected to prior experimental and computational literature, and the authors carefully position their findings within existing knowledge of membrane-associated IDPs.

Weaknesses:

A primary limitation is the heavy reliance on computational modeling. Training for AroMIP is generated using PPM rather than direct experimental measurements, and so the model may primarily reproduce PPM behavior rather than true membrane insertion thermodynamics. Moreover, all simulations use a single lipid composition (POPC:POPS:PIP₂ 70:25:5), but biological membranes vary substantially in cholesterol, cardiolipin, and acidic lipid content. Whether AroMIP's predictions transfer to diverse lipid environments remains untested. The 5% PIP₂ concentration used in the simulations is higher than that of a normal mammalian cell and may therefore overemphasize electrostatic contributions. Applicability beyond short 9-residue motifs is unclear, as longer-range interactions or secondary structure in full-length IDRs could modulate insertion in ways the current model does not capture. This could be considered for future development.

<https://doi.org/10.7554/eLife.111515.1.sa3>

Reviewer #2 (Public review):

Summary:

The paper addresses an interesting problem. The authors develop a method to assess the probability of insertion of aromatic residues in intrinsically disordered regions of proteins, to insert in the interfacial regions of membranes.

Strengths:

- (1) The idea of the article seems very interesting. The problem of membrane association mediated by aromatic residues is definitely worth studying. Aromatic residues, especially Tryptophan (W), but also, albeit to a lesser extent, Phenylalanine (F), and Tyrosine (Y), are well known to partition preferentially to the headgroup region of the lipid bilayer.
- (2) The authors propose to decipher the sequence code for insertion of sequences containing aromatic residues in the membrane employing three types of calculation methods with decreasing order of detail and complexity, but increasing order of efficiency. First, all-atom MD simulations; second, the PPM method (protein positioning in membranes) from Lomize et al (2006), *Protein Sci* 15, 1318; and third, AroMIP, a mathematical model developed by the authors. The results obtained with the different simulations and mathematical methods are internally consistent.

Weaknesses:

- (1) Aromatic residues have been shown to partition preferentially to the headgroup region of the lipid bilayer. Most of the papers on this problem were published in the mid 1990s to early 2000s. Some of the most important papers in this regard are the following: von Heijne, *Annu. Rev. Biophys. Biomol. Struct.* 1994, 23, 167-192; Doyle et al. *Science* 1998, 280, 69-77; Landolt-Marticorena, et al. *J. Mol. Biol.* 1993, 229, 602-608; Killian & von Heijne, *TIBS* 2000, 25, 429-434; Marx & Fleming *J. Am. Chem. Soc.* 2021, 143, 764-772. Strangely enough, none of these articles is cited.
- (2) This is the most important point and the most serious weakness. The authors find that the PPM method is able to reproduce the results from MD simulations, and the AroMIP model is able to perform well in comparison with PPM and MD, after training AroMIP on a large set of IDR sequences (intrinsically disordered protein regions) of the human proteome. The

defining feature of the AroMIP calculation is the recognition of the importance of flanking residues in the membrane-insertion propensity of a sequence containing a central aromatic residue. All this sounds good. However, this is all theoretical. There is no connection to experiment or to any method that draws from experiment. The entire approach relies on the assumption that the MD simulations produce the correct results. There is no proof of the correctness of anything. As one of the greatest physicists of our times, Richard Feynman, wrote, "The test of all knowledge is experiment. Experiment is the sole judge of scientific "truth"."

(3) The drawings in Figures 2 and 3 are incorrect and misleading. The size of the Tryptophan side chain is about 5.5 Å, whereas one-half of the bilayer ("a monolayer") thickness is about 15 Å. But in the figures, the lipid length and the Trp side chain seem about the same size. This is incorrect even in a qualitative sense.

<https://doi.org/10.7554/eLife.111515.1.sa2>

Reviewer #3 (Public review):

Summary:

This is a well-written manuscript that describes three robust and complementary computational approaches to unravel the sequence determinants of membrane insertion, specifically of intrinsically disordered regions (IDRs) containing aromatic-centered insertion motifs.

Strengths:

A robust, multifaceted computational approach employing aromatic-centered model membrane-insertion peptides, which provides critical insights into the determinants of membrane insertion.

Weaknesses:

I only have specific concerns about some of the models used for this purpose.

(1) Membrane composition and lipid shape characteristics: The authors chose to use a model membrane bilayer of a distinct lipid composition, POPC: POPS: PI4,5P2 (70:25:5 molar ratio), for their all-atom simulations of the various model peptides. While this may be pertinent for some of these peptides, it is not for many, such as sequence 2 derived from Drp1, which preferentially binds target conical lipids such as cardiolipin (CL) and phosphatidic acid (PA). The rationale behind using PI4,5P2, which can induce positive membrane curvature when sequestered, versus CL and PA, which both induce negative membrane curvature, is not explained.

(2) Parallel vs. perpendicular peptide orientation of sequence 2 in peripheral Drp1-lipid interactions: On page 11, the authors state that their simulation results of sequence 2 derived from Drp1 "contrasts with a transmembrane orientation proposed by Mahajan et al." However, upon review, a transmembrane orientation for this region has never been proposed anywhere. Drp1 is a peripheral membrane protein that reversibly binds CL- and PA-containing membranes via its intrinsically disordered variable domain containing an aromatic-centered WRG motif. Indeed, the model presented in Figure 9 of Mahajan et al. displays a peripheral and parallel orientation of the transiently helical WRG-containing motif rather than a transmembrane (i.e., across the bilayer) orientation. While the authors can distinguish between a parallel vs. perpendicular orientation of this sequence relative to the plane of the membrane bilayer surface from their simulations, suggesting that previous studies indicated a transmembrane orientation for Drp1 is disingenuous and misleading. The term "transmembrane" should be removed or replaced, as it presents a wrong image.

(3) Mutational analysis of W vs. F in membrane insertion of W-centered insertion motifs and vice versa: The PPM-based workflow suggests that F-centered sequences have the highest membrane insertion properties as opposed to W-centered ones. A W552F mutation in the WRGML sequence of Drp1 was, however, found to impair function. How do the authors rationalize this? A cross-mutational analysis of W vs. F in W-centered motifs and F-centered motifs is warranted.

<https://doi.org/10.7554/eLife.111515.1.sa1>

Author response:

eLife Assessment

This study provides a valuable advance in understanding how disordered proteins interact with cell membranes by identifying the sequence rules that enable aromatic residues to penetrate deeply into the membrane interior. The integration of complementary computational approaches, including molecular simulations, large-scale sequence analysis, and the development of an online prediction server, makes the work potentially impactful for the membrane protein and intrinsically disordered protein communities. The evidence supporting the main conclusions is generally convincing, although its transferability across diverse membrane compositions and its validity as a prediction tool for real protein-membrane systems remain to be further established.

We thank the editors for recognizing our study as a valuable advance. This work lays a solid foundation for future developments to account for diverse membrane compositions and further refinements after additional experimental tests.

Public review:

Reviewer #1:

A primary limitation is the heavy reliance on computational modeling. Training for AroMIP is generated using PPM rather than direct experimental measurements, and so the model may primarily reproduce PPM behavior rather than true membrane insertion thermodynamics. Moreover, all simulations use a single lipid composition (POPC:POPS:PIP₂ 70:25:5), but biological membranes vary substantially in cholesterol, cardiolipin, and acidic lipid content. Whether AroMIP's predictions transfer to diverse lipid environments remains untested. The 5% PIP₂ concentration used in the simulations is higher than that of a normal mammalian cell and may therefore overemphasize electrostatic contributions. Applicability beyond short 9-residue motifs is unclear, as longer-range interactions or secondary structure in full-length IDRs could modulate

insertion in ways the current model does not capture. This could be considered for future development.

The reviewer's point on our reliance on PPM for training, a single lipid composition, and potential effects beyond a 9-residue motif is well taken. Regarding PPM, we chose it as the optimal compromise for high-throughput data. However, we complemented the high-throughput PPM data with experimental data on an initial set of 10 peptides. Moreover, we validate AroMIP on an additional 12 IDRs (intrinsically disordered regions; Table S2). On membrane composition, we now acknowledge the limitation of our work based on a single composition and point to future developments of AroMIP involving membrane-specific parameterization (p. 19, 3rd paragraph). On potential effects beyond a 9-residue motif, we now add justification and note neglected factors for future developments (paragraph running from p. 19-20), as suggested by the reviewer.

Reviewer #2:

(1) Aromatic residues have been shown to partition preferentially to the headgroup region of the lipid bilayer. Most of the papers on this problem were published in the mid 1990s to early 2000s. Some of the most important papers in this regard are the following: von Heijne, Annu. Rev. Biophys. Biomol. Struct. 1994, 23, 167-192; Doyle et al. Science 1998, 280, 69-77; Landolt-Marticorena, et al. J. Mol. Biol. 1993, 229, 602-608; Killian & von Heijne, TIBS 2000, 25, 429-434; Marx & Fleming J. Am. Chem. Soc. 2021, 143, 764-772. Strangely enough, none of these articles is cited.

We have now citations to the Landolt-Marticorena paper and the von Heijne reviews (refs 25-27). The Doyle paper is not particularly relevant. As for the Fleming paper, we cited a 2016 JACS paper (original ref 27; now ref 30) that specifically dealt with aromatic residues.

(2) This is the most important point and the most serious weakness. The authors find that the PPM method is able to reproduce the results from MD simulations, and the AroMIP model is able to perform well in comparison with PPM and MD, after training AroMIP on a large set of IDR sequences (intrinsically disordered protein regions) of the human proteome. The defining feature of the AroMIP calculation is the recognition of the importance of flanking residues in the membrane-insertion propensity of a sequence containing a central aromatic residue. All this sounds good. However, this is all theoretical. There is no connection to experiment or to any method that draws from experiment. The entire approach relies on the assumption that the MD simulations produce the correct results. There is no proof of the correctness of anything. As one of the greatest physicists of our times, Richard Feynman, wrote, "The test of all knowledge is experiment. Experiment is the sole judge of scientific "truth"."

We emphasize that we have presented substantial experimental support for AroMIP. It correctly predicts the membrane insertion status of the initial set of 10 peptides, which were characterized experimentally. In addition, we validated AroMIP on an additional set of 12 IDRs (Table S2), most of which were characterized by experimental techniques including solution and solid-state NMR, fluorescence, H/D exchange, and cryo-EM. Lastly, we now show good correlation between our insertion scores and binding free energies calculated from the scale determined experimentally by White and co-workers (new Figure S10; p. 15, second paragraph).

(3) The drawings in Figures 2 and 3 are incorrect and misleading. The size of the Tryptophan side chain is about 5.5 Å, whereas one-half of the bilayer ("a monolayer") thickness is about 15 Å. But in the figures, the lipid length and the Trp side chain seem about the same size. This is incorrect even in a qualitative sense.

We have now revised these figures.

Reviewer 3:

(1) Membrane composition and lipid shape characteristics: The authors chose to use a model membrane bilayer of a distinct lipid composition, POPC: POPS: PI4,5P2 (70:25:5 molar ratio), for their all-atom simulations of the various model peptides. While this may be pertinent for some of these peptides, it is not for many, such as sequence 2 derived from Drp1, which preferentially binds target conical lipids such as cardiolipin (CL) and phosphatidic acid (PA). The rationale behind using PI4,5P2, which can induce positive membrane curvature when sequestered, versus CL and PA, which both induce negative membrane curvature, is not explained.

We now acknowledge the limitation of our work based on a single composition and point to future developments of AroMIP involving membrane-specific parameterization (p. 19, 3rd paragraph). In this Discussion paragraph, we also speculate that conical lipids, by promoting membrane defects, may facilitate membrane insertion.

(2) Parallel vs. perpendicular peptide orientation of sequence 2 in peripheral Drp1-lipid interactions: On page 11, the authors state that their simulation results of sequence 2 derived from Drp1 "contrasts with a transmembrane orientation proposed by Mahajan et al." However, upon review, a transmembrane orientation for this region has never been proposed anywhere. Drp1 is a peripheral membrane protein that reversibly binds CL- and PA-containing membranes via its intrinsically disordered variable domain containing an aromatic-centered WRG motif. Indeed, the model presented in Figure 9 of Mahajan et al. displays a peripheral and parallel orientation of the transiently helical WRG-containing motif rather than a transmembrane (i.e., across the bilayer) orientation. While the authors can distinguish between a parallel vs. perpendicular orientation of this sequence relative to the plane of the membrane bilayer surface from their simulations, suggesting that previous studies indicated a transmembrane orientation for Drp1 is disingenuous and misleading. The term "transmembrane" should be removed or replaced, as it presents a wrong image.

We have now deleted the sentence mentioning "transmembrane orientation".

(3) Mutational analysis of W vs. F in membrane insertion of W-centered insertion motifs and vice versa: The PPM-based workflow suggests that F-centered sequences have the highest membrane insertion properties as opposed to W-centered ones. A W552F mutation in the WRGML sequence of Drp1 was, however, found to impair function. How do the authors rationalize this? A cross-mutational analysis of W vs. F in W-centered motifs and F-centered motifs is warranted.

AroMIP predicts a membrane insertion propensity of 0.782 for the WRGML sequence and a moderately higher propensity, 0.837, with a W552F mutation. This increase contradicts the experimental observation of a 3.6-fold increase in membrane binding affinity by Mahajan et al. We now speculate that the specific lipid, cardiolipin, as the reason for the discrepancy (p. 19, 3rd paragraph). This discrepancy provides a concrete example for the need to account for membrane composition in future developments.

<https://doi.org/10.7554/eLife.111515.1.sa0>